Exploring the Heterogeneous Impact of Immigrants on Local Labor Market Outcomes

Ruxin Chen
Master of Computational Social Science Program (Social Science Division), University of Chicago

Background

The effect of immigrants on the employment opportunities for native workers has received a great deal of research attention during the last decades. In recent years, as the intensifying of the regional terrorist threats in the Middle East, whether to accept refugees has been at the heart of a growing political debate for many western countries. Moreover, with the slowing pace of economic growth in the developed world, policy makers seem to be concerned about the adverse effect, if any, on employment that immigrants might bring to the native labor market.

The classical model suggests that an inflow of immigrants would always lower the wage of competing local workers. Though the model is well understood, the economic mechanism of the interplay of immigrants and native workers might go beyond the over-simplified demand-supply model. Moreover, the empirical studies regarding the association between immigrants and local labor market outcomes do not reach a consensus in existing literature.

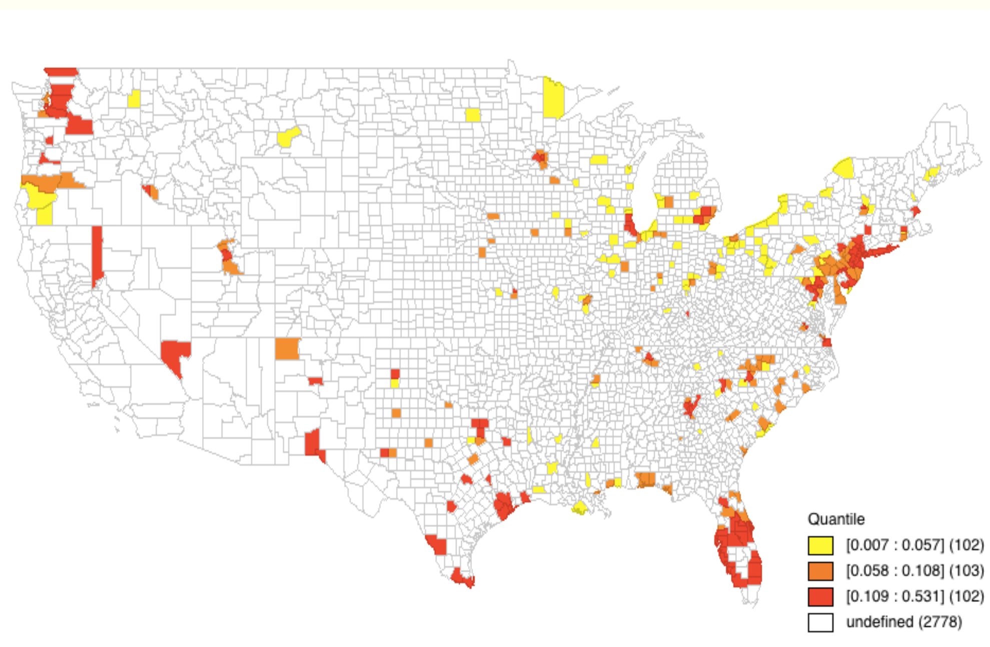
Contribution To the Literature

It is reasonable to speculate the factors are masked by the heterogeneity of the population. Hence, my project deploys a regression-tree based method to identify the demographical heterogeneity in effect of immigration on native labor market outcomes on different subgroups of the population. Two dimensions are investigated specially: the employment rate and average wage and salary income. This study is expected to provide some policy implications under the current political landscape of exclusionism/protectionism.

Data Description

We initially requested a data set consists of more than 3,800,000 observations during 2006 – 2016 from IPUMS USA. Variables include state, county, metropolitan status, age, sex, marital status, race, ethnicity, year of immigration, language spoken, education attainment, labor force status, employment status, occupation, class of worker, usual hours worked per week and wage and salary income. The data contain 423 counties that have a wide range of spatial location.

Proportion of Immigrants in Data



Methodology

The machine learning method I applied is based on the method proposed by Athey and Imbens (2006), which uses a regression tree to find the partition of the population according to covariates. Most machine learning technique cannot be used directly for constructing confidence interval since the models are "adaptive" and the conventional asymptotic properties cannot be achieved. They propose a two-steps procedure to overcome this problem, called "honest" approach, which splits the training sample into two parts for constructing the tree and estimating treatment effect within each leaf of the tree. The independence of the two steps guarantees the validity of the estimation.

Model Specification and Preliminary Regression Results

Before building the decision tree, we construct the OLS and Fixed Effect Model :

OLS: $y_i = \beta_0 + \beta_1 Proportion\ of\ Immigrants_{i,} + X_i + u_i$

Fixed Effect: $y_{i,t} = \beta_0 + \beta_1 Proportion\ of\ Immigrants_{i,t} + X_{i,t} + \alpha_i + \eta_t + u_{i,t}$

where the unit of observation is county, y denotes the outcome of interest, rate of unemployment and the average wage income, X is the collection of control variables including the average age for the county, proportion of female, proportion of white/black population, proportion of Hispanics, and the education attainment level. α_i denotes the state fixed effect and η_t is the year fixed effect.

Table 1: Preliminary Regression Results for Immigrants and Unemployment Rate

	(1)	(2)	(3)
	OLS	Fixed Effect	Fixed Effect
Proportion of Immigrants	0.0897***	0.0739***	0.0052
	(0.01)	(0.01)	(0.01)
Constant	0.1882	0.1640	0.2655***
	(0.10)	(0.08)	(0.07)
Year Fixed Effect	No	Yes	Yes
State Fixed Effect	No	No	Yes
R-squared	0.4170	0.5952	0.7339
N. of cases	4450	4450	4450

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Preliminary Regression Results for Immigrants and Average Wage Income

	(1)	(2)	(3)
	OLS	Fixed Effect	Fixed Effect
Proportion of Immigrants	15132.29***	19611.43***	17277.98***
	(1158.14)	(1262.44)	(1445.59)
Constant	-2830.91	-47928.35*	-44704.04*
	(19118.59)	(19596.96)	(17479.36)
Year Fixed Effect	No	Yes	Yes
State Fixed Effect	No	No	Yes
R-squared	0.8105	0.8201	0.8683
N. of cases	4450	4450	4450

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From above tables, we can see that after the inclusion of year and state fixed effect, the proportion of immigrants are insignificant to the unemployment rate, however, it has a significant positive impact on the average wage income.

Step 1:Regression Tree From Training Sample

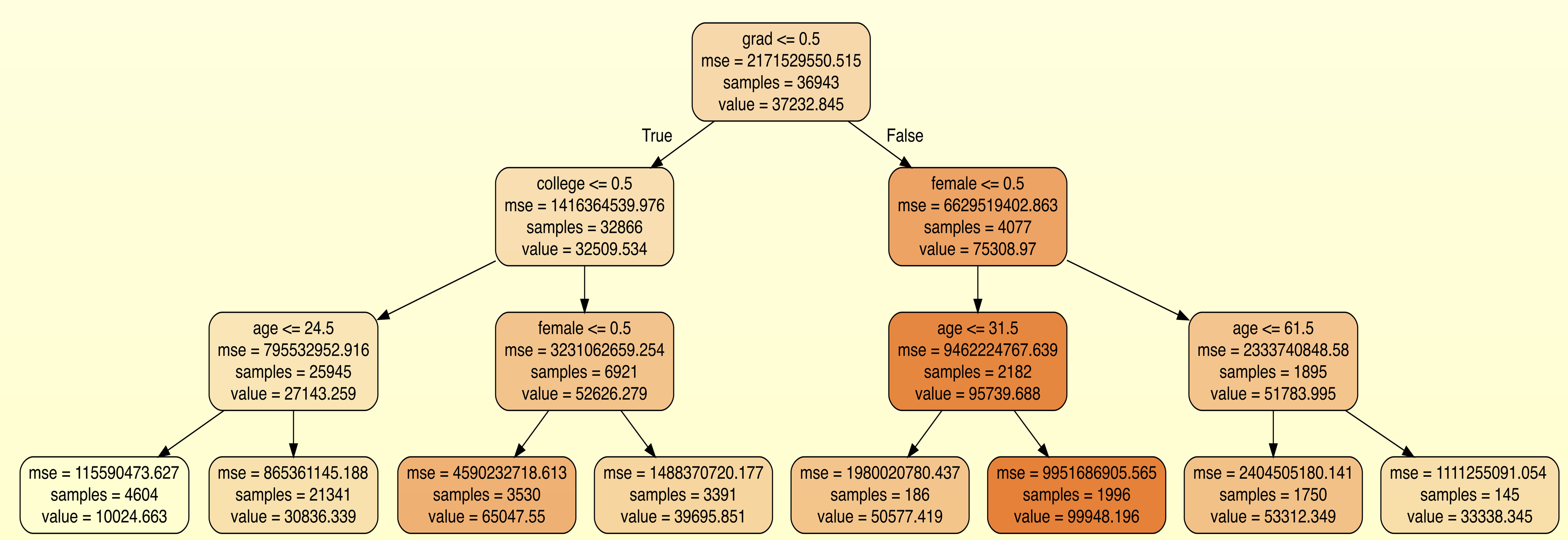
The Conventional Average Regression Tree (CART) algorithm recursively partitions the observations of the training sample. For each leaf, the algorithm evaluates all candidates splits of that leaf using a "splitting" criterion that is referred as the "in-sample" goodness-of-fit criterion:

$$min_{j,c}\ MSE = min_{j,c}[min_c \sum_{x_i \in R_1(j,c)} (y_i - c_1)^2 + min_c \sum_{x_i \in R_2(j,c)} (y_i - c_2)^2]$$

where, consider a splitting variable j and split point s,

$$R_1(j,s) = \{X | X_j \le s\} \text{ and } R_2(j,s) = \{X | X_j > s\}$$

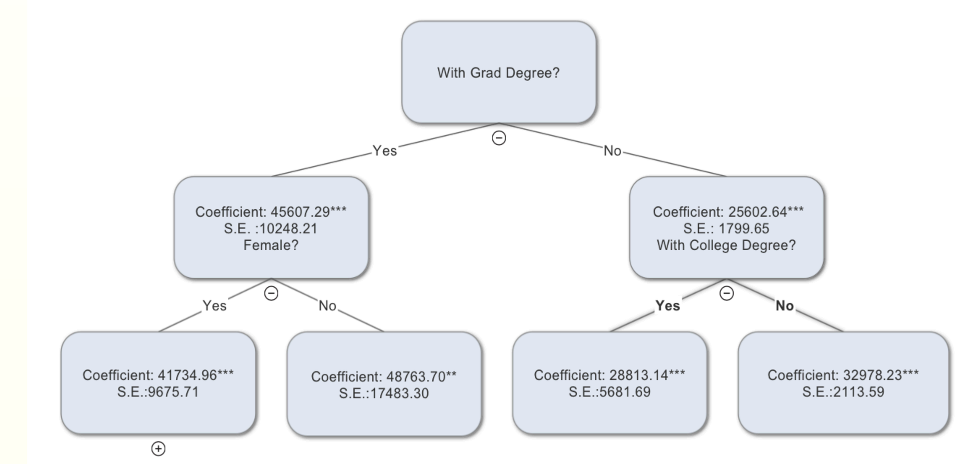
According to the "honest" method, we need to make a adjustment of the criterion MSE by subtracting two times of the within-leaf sample variances, $\frac{2}{N^{tr}} \sum_{l \in \Pi} s_{tr}^2(l)$. Then we get a regression tree as follows:



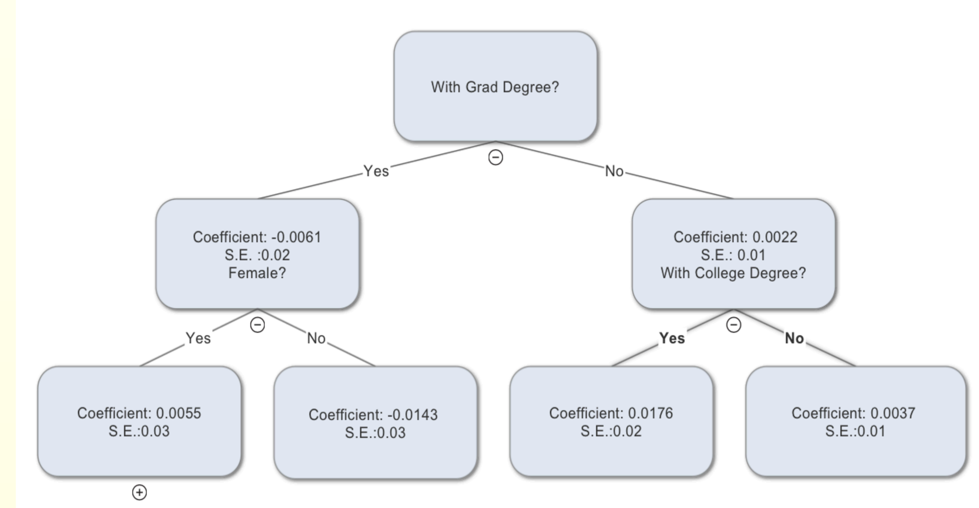
Step 2: Estimation & Results

In the second step, we use another half of data to do the estimation, the results are presented here:

- In general, the wage income will be increased as the proportion of immigrants increases. The magnitude of the effect is found to be larger for individuals with advanced degree.



- The probability of being unemployed is estimated by a logit regression. From the graph below we can see that though the signs of coefficient are different, the effect of immigrants are insignificant across all subpopulations.



Limitations

Though the econometric model incorporates year and state fixed effects to take account for the unobserved endogeneity, I would be cautious to claim the findings in the paper are causal. Future studies might focus on figuring out an identification strategy for this topic.

References

Athey, Susan, and Guido Imbens. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences* 113, no. 27 (2016): 7353-7360.